

What is claimed is:

- 1 1. A method for dynamically identifying clusters of related data comprising:
2 launching a probe from a first position in an M-dimensional space, said M-
3 dimensional space having a plurality of data points, each of said plurality of data points
4 associated with a data record, each data record having at least M number of data fields;
5 determining a new position for said probe in said M-dimensional space based on a
6 current position of said probe relative to at least a portion of said plurality of data points in
7 said M-dimensional space;
8 moving said probe from said current position to said new position;
9 repeating said determining a new position for said probe until said new position and
10 said current position are approximately a same position;
11 dynamically identifying a cluster upon determining said same position in said M-
12 dimensional space.
- 1 2. The method of claim 1, further comprising launching another probe from another
2 position in said M-dimensional space to initiate identification of another cluster in said M-
3 dimensional space.
- 1 3. The method of claim 1, wherein said launching another probe from another position
2 comprises randomly determining said another position in said M-dimensional space.
- 1 4. The method of claim 1, wherein said launching another probe from another position
2 comprises determining said another position in said M-dimensional space outside a
3 predetermined proximity from a previous probe trajectory.
- 1 5. The method of claim 1, wherein said launching another probe from another position
2 comprises determining said another position in said M-dimensional space beyond a
3 predetermined proximity from said identified cluster.
- 1 6. The method of claim 1, wherein said launching another probe from another position
2 comprises determining said another position as one of said plurality of data points.

1 7. The method of claim 1, wherein said launching another probe from another position
2 comprises determining said another position in said M-dimensional space as one of said
3 plurality of data points that is outside of a predetermined proximity from a previous probe
4 trajectory and beyond a predetermined proximity from said identified cluster.

1 8. The method of claim 1, wherein said determining a new position for said probe
2 comprises applying a localized force function to said probe.

1 9. The method of claim 8, wherein said determining a new position for said probe
2 comprises applying a localized coulomb force function to said probe.

1 10. The method of claim 8, wherein said determining a new position for said probe
2 comprises applying a localized force function to said probe, said force function based on a
3 radial distance between said probe and each of said plurality of data points.

1 11. The method of claim 8, wherein said determining a new position for said probe
2 comprises applying a potential function to said probe.

1 12. The method of claim 8, wherein said determining a new position for said probe
2 comprises applying a potential function to said probe, said potential function based on at
3 least one of a weight function and a quadratic function.

1 13. The method of claim 12, wherein said applying a potential function to said probe
2 comprises applying a product of a weight function and a quadratic function.

1 14. The method of claim 13, wherein said applying a potential function to said probe
2 comprises applying a potential function of the form $V=R^2 \cdot \exp(-R^2/2 \sigma^2)$, where "V" is the
3 potential between said probe and one of said plurality of data points, "R" is the distance in
4 said M-dimensional space between said probe and said one of said plurality of data points,
5 and " σ^2 " is an estimate of noise variance associated with said plurality of data points.

1 15. The method of claim 11, further comprising minimizing a sum of said potential
2 functions applied to each of said at least a portion of said plurality of data points.

0998746-10901

1 16. The method of claim 15, wherein said minimizing a sum of said potential functions
2 comprises minimizing a sum of said potential functions applied to each of said plurality of
3 data points.

1 17. A method for dynamically identifying a number of clusters of related data from a
2 plurality of data records each having a plurality of data fields, the data represented as N
3 data points in an M-dimensional space where M is less than or equal to a number of the
4 plurality of data fields and N is less than or equal to a number of the plurality of data
5 records, the method comprising:

6 initializing a current position of a data probe as a first position in the M-dimensional
7 space;

8 determining a new position for said data probe in the M-dimensional space based on
9 a similarity between said data probe as indicated by said current position and at least a
10 portion of the N data points in the M-dimensional space;

11 adjusting said current position of said data probe to said new position;

12 repeating said determining a new position and said adjusting said current position
13 until said new position and said current position are approximately a same position; and

14 once said new position and said current position are approximately said same
15 position, incrementing a count of the number of clusters of related data.

1 18. The method of claim 17, further comprising:

2 reinitializing a current position of said data probe as a second position in the M-
3 dimensional space, said second position different from said first position.

1 19. The method of claim 18, further comprising:

2 repeating said determining a new position and said adjusting said current position
3 until said new position and said current position are approximately a second same position.

1 20. The method of claim 19, further comprising:

2 if said second same position is a unique same position, then incrementing said count
3 of the number of clusters of related data.

09986746-10901

1 25. The method of claim 17, wherein said determining a new position for said data
2 probe in said M-dimensional space comprises determining a relative distance between said
3 data probe and each of said at least a portion of the N data points.